REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 28-02-2006	**2. REPORT TYPE** Final Report	**3. DATES COVERED** *(From – To)* 18-Mar-02 - 06-Mar-06

4. TITLE AND SUBTITLE

Development of Mathematical Models of Immune Networks Intended for Information Security Assurance

5a. CONTRACT NUMBER
ISTC Registration No: 2200p

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Alexander O Tarakanov

5d. PROJECT NUMBER

5d. TASK NUMBER

5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

St. Petersburg Institute For Informatics & Automation of the Russian Academy of Sciences
39, 14th Liniya, St.Petersburg, 199178, Russia
Saint Petersburg
Russia

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

EOARD
PSC 821 BOX 14
FPO 09421-0014

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
ISTC 01-7007

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The natural immune system is considered by many specialists as a "second brain of vertebrates". In fact, the immune system possesses all the main features of Artificial Intelligence (AI) systems: (1) memory, (2) learning capability, (3) capability to recognize self and non-self, and (4) decision-making capability, that is, the immune system must decide how to treat all macromolecules it encounters even if such molecules are foreign and have never existed before. Of special interest to computer science is the theory of immune networks which describes interactions between immune system specific proteins (antibodies) and foreign macromolecules (antigens). The existence of such immune networks has been established experimentally by molecular immunology which has detected and described the antibody-antigen interaction. Based on the biological principles of the immune system, the field of Artificial Immune Systems (AISs) has been established. It hopes to offer powerful and robust information processing capabilities for solving complex problems. For example, AISs may provide improved techniques to detect and mitigate modern computer network vulnerabilities to intrusions from computer viruses, unauthorized access or other forms of data corruption. Like other modern computer science techniques such as Artificial Neural Networks (ANNs) or Intelligent Agents, AISs can learn new information, recall previously learned information, and perform pattern recognition in a highly decentralized fashion. However, AISs based on natural immune networks differ remarkably from ANNs, intelligent agents, genetic algorithms, and cellular automata in their ability to recognize self and non-self and their highly specific activity. AISs have already been applied to several specific problems including information security, fault detection, robotic control, and others.

15. SUBJECT TERMS

EOARD, Mathematical And Computer Sciences, Computer Programming and Software

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18, NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** PAUL LOSIEWICZ, Ph. D.
a. REPORT UNCLAS	**b. ABSTRACT** UNCLAS	**c. THIS PAGE** UNCLAS	UL	33	**19b. TELEPHONE NUMBER** *(Include area code)* +44 20 7514 4474

Final
Project Activities Report
# of ISTC 2200p

Development of Mathematical Models of Immune Networks Intended for Information Security Assurance

(From 1 February 2002 to 31 January 2006 for 48 months)

Alexander Olegovich Tarakanov
(Project Manager)
St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences

February 2006

This work was supported financially by European Office of Aerospace Research and Development (EOARD) and performed under the contract to the International Science and Technology Center (ISTC), Moscow.

Development of Mathematical Models of Immune Networks Intended for Information Security Assurance (From 1 February 2002 to 31 January 2006 for 48 months)

Alexander Olegovich Tarakanov
(Project Manager)
St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences *

The objective of this project is to develop a novel approach to information assurance (IA) based on a rigorous mathematical notion of formal immune network (FIN).

A special kind of FIN (cFIN) intended for IA has been developed and implemented in so-called immunochip emulator. This software emulator has been tested on data simulating intrusions in a typical computer network (UCI KDD archive). Training time over a training set of about 51000 network connection records is about 60s (AMD 1.5GHz). Fine-tuning of the emulator reduces the number of storing patterns and thus the recognition time per pattern by 60 times at least. The emulator correctly recognizes all intrusions in the training set by 16ms per record.

The comparison with neural computing and genetic algorithms over nother real-life tasks of pattern recognition (in ecology and laser physics) also demonstrates that the performance of FIN (training time and accuracy) is unachievable for other approaches of computational intelligence.

A hardware implementation of FIN has been proposed based on digital signal processor of super Harvard architecture (DSP SHARC).

Keywords (about 10 words): Immunocomputing, Information Assurance, Formal Immune Network, Immunochip

*199178, St. Petersburg, 14-line, 39, Russia
Tel: +7 (812) 328 4450, Fax: +7 (812) 328 0685, E-mail: spiiran@iias.spb.su

The work has been performed by the following Institute and Collaborator:

1. Participated Institute:
Saint-Petersburg Institute for Informatics and Automation (SPIIRAS)
14-line, 39, St. Petersburg, 199178, Russia
Tel: +7 (812) 328 4450, Fax: +7 (812) 328 0685, E-mail: spiiran@iias.spb.su

2. Foreign Collaborator:
European Office of Aerospace Research and Development (EOARD)
223-231 Old Marylebone Road, London NW1 5TH United Kingdom
Roberta J. Hauck, USAF EOARD Program Coordinator
Tel: +44 207 514 4354, Fax: +44 207 514 4960, E-mail: roberta.hauck @london.af.mil
Paul Losiewicz, Ph. D., Chief, Information/C41
Tel: +44 207 514 4474, Fax: +44 207 514 4960, E-mail: paul.losiewicz@london.af.mil

Objectives of the Project

The objective of this project is to develop a novel approach to information assurance (IA) based on a rigorous mathematical notion of formal immune network (FIN).

A special kind of FIN for IA is expected to be developed with the following capabilities:

- FIN as a system of computational intelligence;
- FIN as an alternative to the wide spread artificial neural networks and intelligent agents;
- FIN as a mathematical and algorithmic basis for hardware implementation of IA issues in a special 'immunochip'.

Scope of Work and Technical Approach

The scope of work and technical approach of the project are as follows:

- Qualitative description of the biological immune networks from the viewpoint of IA.
- Mathematical description of FIN.
- Mathematical description of the special kind of FIN intended for IA.
- Computer implementation of FIN.
- Computer emulator of the immunochip.
- Proposals on hardware implementation of the immunochip.

The 4th year extension of the project has been proposed. It aims to develop a rigorous mathematical model of immune modulation by cytokines (messenger proteins, which play a central role in regulation of immune response) and a novel notion of cytokine FIN (cFIN) intended for IA applications.

An impact of such extension is twofold:
1) The enhanced performance of the software emulator of the immunochip for intrusion detection;
2) The advanced architecture of the immunochip.

The goal of the project is accomplished in three phases through the following research tasks:

Phase I

Task 1. Qualitative description of the immune networks.
Deliverable 1 (after 3 months, Report): Principles of information processing by biological immune networks from the viewpoint of information security.
Task 2. Mathematical description of FIN.
Deliverable 2 (after 6 months, Report): Mathematical notion of FIN, mathematical properties of FIN, main theorems describing the behavior of FIN.
Task 3. Mathematical description of the special kind of FIN intended for information security assurance.
Deliverable 3 (after 9 months, Report): Mathematical description of FIN intended for pattern recognition with information security data.
Task 4. Software implementation of FIN.
Deliverable 4 (after 12 months, Report and Software): Algorithms of pattern recognition by FIN plus demo version of the software.

Phase II

Task 5. Software emulation of the immune chip.
Deliverable 5 (after 15 months, Report): Architecture of the immune chip emulator.
Deliverable 6 (after 18 months, Report): Algorithms of the emulator.
Deliverable 7 (after 21 months, Report and Publication): Software implementation of the algorithms.
Deliverable 8 (after 24 months, Report and Software): User interface of the emulator plus demo version of the software.

Phase III

Task 6. Deployment, testing and fine-tuning of the immune chip emulator on the real word data of IA.
Task 6.1. Deliverable 9 (after 27 months, Report): Description of data and procedures for testing the emulator.
Task 6.2. Deliverable 10 (after 30 months, Report): Deliverable 10: Test results and resume.
Task 6.3. Deliverable 11 (after 33 months, Report and Publication): Fine-tuning of the emulator.
Task 6.4. Deliverable 12 (after 36 months, Report and Software): Proposal on a hardware implementation of the emulator plus demo version of the software.

Task 6.5. Mathematical models of the immune modulation by cytokines.

Deliverable 13 (after 39 months, Report) Development of mathematical models of the immune modulation of FIN by cytokines.

Deliverable 14 (after 42 months, Report) Development of the special kind of cytokine FIN (cFIN) intended for IA.

Task 6.6. The application of cFIN to intrusion detection.

Deliverable 15 (after 45 months, Report): Implementation of cFIN in the software emulator of the immunochip.

Deliverable 16 (after 48 months, Final Report, Publication, Software): Final reports plus demo version of the software.

Summary of Project Technical Report

Method

Cytokines (messenger proteins) are a group of biologically active mediator molecules that provide the intercellular interactions within the immune system. They are the central regulators of leukocyte growth and differentiation, being produced by a wide variety of cell types, targeting various cell subsets and exhibiting numerous biological activities.

Up to now more than 100 different human cytokines are identified. An increasing volume of experimental data suggests that cytokines play one of the central roles in the immune regulation as well as in the neuro-immune-endocrine modulation.

Recent developments show that cytokines induce apoptosis (programmed cell death) in cancer cells. The induction of apoptosis is associated with a dose-dependent inhibition of cancer cell division, and this activity has been demonstrated for a wide range of cancer types including bladder, breast, leukemia, melanoma, ovarian and prostate.

Apoptosis is a natural mechanism by which cells "commit suicide" when they have outlived their purpose, become defective, or have aged. Apoptosis prevents cells from accumulating and forming tumors. Understanding of the control of apoptosis in normal and malignant cells will help to improve the diagnosis and treatment of malignancies. The goal of many treatments, including chemotherapies is to induce malignant cells to undergo apoptosis. Current data also suggests that a cytokine may function as a dual-acting cytokine in which its normal physiological functions may be related to specific aspects of the immune system and over-expression culminates in cancer-specific apoptosis.

On the other hand, immunological approach looks rather constructive as a basis for a new kind of computing [1]. In such background, this project develops a rigorous mathematical model of immune network with the cytokine controlled apoptosis and immunization. A software implementation of the model has been applied to the task of intrusion detection in a local area network (LAN).

Mathematical model

Cytokine formal immune network

Definition 1. Cell is a pair *V = (c, P)*, where "cytokine" *c* is natural number $c \in N$, whereas $P = (p_1,..., p_q)$ is a point of *q*-dimensional Euclidian space: $P \in R^q$, and P lies within unit cube: $\max\{|p_1|,...,|p_q|\} \le 1$.

Let distance ("affinity") $d_{ij} = d(V_i, V_j)$ between cells V_i and V_j be as follows:

$$d_{ij} = \max\left\{\left|(p_1)_i - (p_1)_j\right|,...,\left|(p_q)_i - (p_q)_j\right|\right\}. \qquad (1)$$

Fix some finite non-empty set of cells ("innate immunity") $W_0 = (V_1,...,V_m)$ with non-zero distance between cells: $d_{ij} \ne 0$, $\forall i, j : i \ne j$.

Definition 2. Cytokine formal immune network (cFIN) is a set of cells: $W \subseteq W_0$.

Definition 3. Cell V_i recognizes cell V_k if the following conditions are satisfied: $c_i = c_k$, $d_{ik} < h$, $d_{ik} < d_{ij}$, $\forall V_j \in W$, $j \ne i$, $k \ne j$, where $h \ge 0$ is given "threshold of affinity".

Let us define the behavior ("maturation") of cFIN by the following two rules.

Rule 1 (Apoptosis). If cell $V_i \in W$ recognizes cell $V_k \in W$ then remove V_i from cFIN.

Rule 2 (Auto-Immunization). If cell $V_k \in W$ is nearest to cell $V_i \in W_0 \setminus W$ among all cells of cFIN: $d_{ik} < d_{ij}$, $\forall V_j \in W$, whereas $c_i \ne c_k$, then add V_i to cFIN.

Let W_A be cFIN as a consequent of application of apoptosis to all cells of W_0. Let W_I be cFIN as a consequence of auto-immunization of all cells of W_A by all cells of W_0. Note that the resulting sets W_A and W_I depend on the ordering of cells in W_0. Further it will be assumed that the ordering is given.

Mathematical properties of cFIN

It is obvious that neither the result of apoptosis W_A nor the result of auto-immunization W_I can overcome W_0 for any innate immunity: $W_A \subseteq W_0$, $W_I \subseteq W_0$, $\forall W_0$. Consider more important and less evident properties of cFIN.

Proposition 1. For any innate immunity W_0 there exists threshold of affinity h_0 such that apoptosis does not change W_0 for any *h* less than h_0: $W_A = W_0$, $\forall h < h_0$.

Proposition 2. For any innate immunity W_0 there exists threshold of affinity h_1 such that consequence of apoptosis and auto-immunization $W_1 = W_I(h_1)$ provides the minimal number of cells $|W_1|$ for given W_0 and any *h*: $|W_1| \leq |W_I(h)|$, $\forall h$, $\forall W_I \subseteq W_0$.

The proofs of Proposition 1 and Proposition 2 can be found in Technical Report as well as in our paper [14].

Application of cFIN to pattern recognition

Let "epitope" ("antigenic determinant") be any point $P = (p_1,..., p_q)$ of *q*-dimensional Euclidian space: $P \in R^q$. Note that any cell of cFIN also contains an epitope, according to Definition 1.

Definition 4. Cell V_i recognizes epitope *P* by assigning him class c_i if the distance $d(V_i, P)$ between the cell and the epitope is minimal among all cells of cFIN: $d(V_i, P) = \min\{d(V_j, P)\}$, $\forall V_j \in W$.

Let pattern be any *n*-dimensional column-vector $Z = [z_1,..., z_n]'$, where $z_1,..., z_n$ are real values and (') is symbol of matrix transposing. Let pattern recognition be mapping of the pattern to an epitope: $Z \rightarrow P \in R^q$, and recognition of the epitope by the class of the nearest cell of cFIN. Let $A_1,..., A_m$ be *n*-dimensional training patterns with known classes $c_1,..., c_m$. Let $A = [A_1,..., A_m]'$ be training matrix of dimension $m \times n$. Consider singular value decomposition (SVD: see, e.g., [1]) of this matrix:

$$A = s_1 Y_1 X_1^{'} + s_2 Y_2 X_2^{'} + s_3 Y_3 X_3^{'} + ... + s_r Y_r X_r^{'} ,$$

where *r* is the rank of matrix *A*, s_k are singular values and Y_k, X_k are left and right singular vectors with the following properties: $Y_k^{'} Y_k = 1$, $X_k^{'} X_k = 1$, $Y_k^{'} Y_i = 0$, $X_k^{'} X_i = 0$, $i \neq k$, $k = 1,..., r$, $s_{k-1} \geq s_k$, $k > 1$.

Consider the following mapping of any *n*-dimensional pattern Z to epitope P:

$$p_k = \frac{1}{s_k} Z' X_k , \; k = 1,..., q , \; q \leq r . \qquad (2)$$

Note that formulas (2) can be treated as "binding energies" between "formal proteins" *Z* ("antigens") and X_k ("antibodies"), according to [1]. Note also, that any epitope obtained by application of formulas (2) to any training pattern lies within unit cube (see Definition 1), according to the above properties of singular vectors.

Software implementation

General description (in a pseudocode) of the cFIN approach to pattern recognition is as follows:

```
Training
{
    1st stage training // map data to cFIN ("antigen processing")
    {
        Get training patterns;
        Form training matrix;
        Compute SVD of the training matrix; // Singular Value Decomposition
        Store n singular values // "binding energies"
        Store n right singular vectors; // "antibody-probes"
        Store left singular vectors; // cells of cFIN
    }
    2nd stage training // compress data by cFIN's "maturation"
    { // compute consecutively for all cells of cFIN:
        Apoptosis;
        Auto-Immunization;
    }
}
```

```
Recognition
{
    Get pattern; // "antigen"
    Map the pattern to cFIN;
    Find nearest cell of cFIN;
    Assign class of the nearest cell to the pattern;
}
```

This algorithm has been implemented in a version of the immunochip emulator (version 6.7) using Visual C++ with build in assembler code of the cytokine affinity function (1) for three-dimensional (3D) Euclidian space ($q = 3$) and OpenGL tools for 3D visualization. Screenshot of the emulator is shown in Fig. 1.

Results

This cFIN approach has successfully been developed, implemented, and tested as the software emulator of the immunochip.

Two data files from KDD archive (Bay S.D. The UCI KDD Archive [http://kdd.ics.uci.edu]. Irvine, CA: University of California, Dept. of Information and Computer Science, 1999) have been used to test the emulator:

- File 1: kddcup_data_10_percent_gz.htm (7.7 MB);
- File 2: kddcup_newtestdata_10_percent_unlabeled_gz.htm (44 MB).

File 1 is the training data file. It contains 51608 network connection records. Any record (file string) has the following format, where parameters 2, 3, 4, 42 are symbolic, while other 38 parameters are numerical (real values):

```
1) duration, 2) protocol_type, 3) service, 4) flag, 5) src_bytes,
6) dst_bytes, 7) land, 8) wrong_fragment, 9) urgent, 10) hot,
11) num_failed_logins, 12) logged_in, 13) num_compromised,
14) root_shell, 15) su_attempted, 16) num_root, 17) num_file_creations, 18) num_shells,
19) num_access_files, 20) num_outbound_cmds,
21) is_host_login, 22) is_guest_login, 23) count, 24) srv_count,
25) serror_rate, 26) srv_serror_rate, 27) rerror_rate,
28) srv_rerror_rate, 29) same_srv_rate, 30) diff_srv_rate,
31) srv_diff_host_rate, 32) dst_host_count, 33) dst_host_srv_count,
34) dst_host_same_srv_rate, 35) dst_host_diff_srv_rate,
36) dst_host_same_src_port_rate, 37) dst_host_srv_diff_host_rate,
38) dst_host_serror_rate, 39) dst_host_srv_serror_rate,
40) dst_host_rerror_rate, 41) dst_host_srv_rerror_rate, 42) attack_type.
```

For example, two records (# 1 and # 745) of File 1 are as follows:

```
0,tcp,http,SF,181,5450,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,
0.00,0.00,1.00,0.00,0.00,9,9,1.00,0.00,0.11,0.00,0.00,0.00,0.00,0.00, normal.
184,tcp,telnet,SF,1511,2957,0,0,0,3,0,1,2,1,0,0,1,0,0,0,0,0,1,1,0.00,
0.00,0.00,0.00,1.00,0.00,0.00,1,3,1.00,0.00,1.00,0.67,0.00,0.00,0.00,
0.00, buffer_overflow.
```

File 1.1 has also been prepared with the same 51608 records of the same format just without the last parameter 42) attack_type.

File 2 contains 311079 records of the same format as in File 1.1.

File 1.1 and File 2 are the test data files.

Note that KDD archive does not indicate the correct types of attack for none of the records of File 2. The only available information on possible attacks is gathered in Tab. 1 (column 'Code' is the emulator's code of attack). Nevertheless, File 2 has been used to test whether the emulator is able to detect unknown intrusions, which had not been presented in the training data of File 1.

The results of training the emulator by File 1 are shown in Fig.1, where right-hand screen represents the initial population of cFIN after SVD (Start cells: $|W_0| = 51608$), while left-hand screen shows cFIN after apoptosis and immunization ($h_1 = 0.5$, $|W_1| = 783$). Total training time (for AMD 1.5GHz) is 62 seconds including 8s for the 1st stage (SVD) and 54 s for the 2nd stage (apoptosis and auto-immunization).

During the recognition of the records of File 1.1 and File 2, the emulator writes test results into the output file in the format: Record # - attack_type. For example, four records (## 744-747) with test results for File 1.1 are as follows (see also Tab. 2):

```
744 - normal.
745 - buffer_overflow. !!!
746 - buffer_overflow. !!!
747 - normal.
```

The emulator also shows on-line projection of any pattern to 3D cFIN (see bold skew cross in both screens) and write the recognition result on the bottom panel (see "Class: back !!!").

Test results in Tab. 2 correspond completely to the correct attack types (parameter 42) of File 1.

Another test has been performed over File 2 to check whether the emulator is able to detect unknown intrusions, which had not been presented in the training data of File 1. The intrusion is treated as unknown if the projection of corresponding pattern to cFIN lies outside of the unit cube (according to Definition 1). The emulator has recognized 13 unknown intrusions as the following records ## of File 2:

417, 12674, 97891, 139795, 170498, 176201, 177958, 232570, 236975, 296561, 296657, 96796, 297658.

According to Tab. 1, any unknown intrusion can correspond to one of the following types of attack that had not been presented in the training data:

apache2, guess_passwd, multihop, named, saint, sendmail, snmpgetattack, udpstorm, xlock, xsnoop.

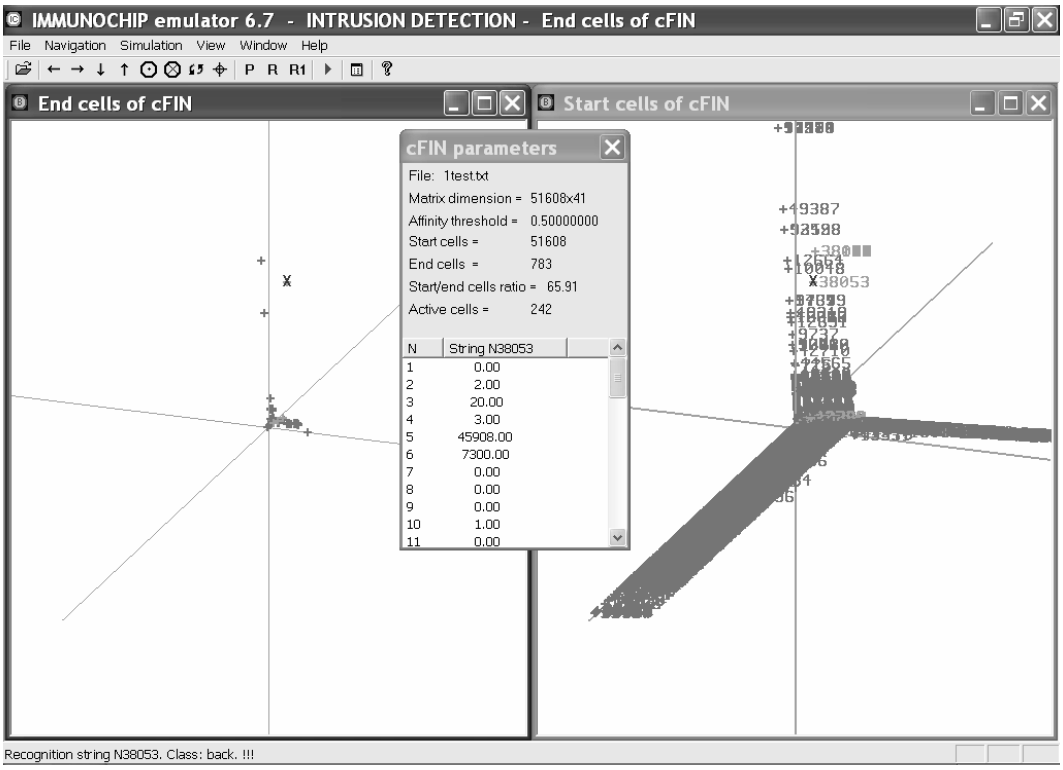
The recognition time per record is 15.7 ms for both tests of File 1.1 and File 2. This time includes not only computations but mainly reading the record from test file, visualization of the recognition result (cFIN's projection of the pattern) in both screens of the emulator and writing the result into output file.

Table 1. Attack types

Code	Attack type	File 1	File 2	Code	Attack type	File 1	File 2
0	normal	+	+				
1	apache2		+	16	pod	+	+
2	back	+		17	portsweep	+	+
3	buffer_overflow	+	+	18	rootkit	+	
4	ftp_write			19	saint		+
5	guess_passwd		+	20	satan	+	
6	imap			21	sendmail		+
7	ipsweep	+	+	22	smurf	+	
8	land	+		23	snmpgetattack		+
9	loadmodule			24	spy		
10	multihop		+	25	teardrop	+	
11	named		+	26	udpstorm		+
12	neptune	+		27	warezclient		
13	nmap			28	warezmaster		
14	perl			29	xlock		+
15	phf	+	+	30	xsnoop		+

Table 2. Test results for File 1.1

Records ##	attack_type	Records ##	attack_type
745-746	Buffer_overflow	38036-38051	ipsweep
3095-7373	Smurf	38052-38151	back
9520-9523	Buffer_overflow	38302-38311	ipsweep
9590-9591	rootkit	42498-42519	ipsweep
9928-10007	neptune	42548-42567	ipsweep
10072	Satan	42593-42594	ipsweep
10320	phf	42706-42708	ipsweep
13340-13519	portsweep	42730-42761	ipsweep
13569	land	42762-42770	buffer_overflow
13845-13864	pod	42771-42772	land
16326-16327	pod	42773-43385	neptune
17446-37902	neptune	44451-44470	neptune
37929-37939	ipsweep	44800-48452	smurf
37959-37963	ipsweep	48453-48552	teadrop
38005-38012	ipsweep	All other	normal



cFIN parameters

File: 1test.txt

Matrix dimension = 51608x41

Affinity threshold = 0.50000000

Start cells = 51608

End cells = 783

Start/end cells ratio = 65.91

Active cells = 242

N	String N38053
1	0.00
2	2.00
3	20.00
4	3.00
5	45908.00
6	7300.00
7	0.00
8	0.00
9	0.00
10	1.00
11	0.00

Recognition string N38053. Class: back. !!!

Fig. 1. Intrusion detection by cFIN: "Antigen" (String 38053 of File 1.1) is mapped to cFIN (skew cross) and recognized by the "cytokine" type of the nearest cell of cFIN (Class: back !!!)

Conclusion

The obtained results suggest that training time and accuracy of the model are beyond the possibilities of artificial neural networks and genetic algorithms [12-15].

According to test results, cFIN reduces the storing patterns by 65.9 times using apoptosis and auto-immunization without any loss of accuracy of recognition. Although this increases the training time (from 8 seconds to 1 minute for AMD 1.5 GHz), nevertheless, more important is the decrease of the recognition time at least by 60 times per pattern by decreasing number of the stored cells of cFIN to be compared with recognizing pattern.

It is worth noting that so good performance of cFIN (error-free recognition with rather low training time) on the data of real-life dimension looks unobtainable for main competitors in the field of computational intelligence like artificial neural networks (ANN) and genetic algorithms (GA). According to the comparison in [12] and [13], cFIN trains by at least 40 times faster and recognizes by at least 2 times correctly than ANN and GA on the tasks of environmental monitoring and laser physics. These tasks have rather low dimension: 17×23×6 for ecological atlas and 19×5 for laser diode. Such drawbacks of ANN and GA become especially inadmissible for the task of intrusion detection with rather high dimension 51608×41 and more.

It is also worth noting that cFIN differs essentially from the negative selection algorithm (NSA). Actually, NSA aims to provide a set of detectors for self-nonself discrimination, whereas cFIN guarantees a minimal set of "cells" for the correct recognition of any number of classes based on "cytokines". Apparently, this makes cFIN advantageous not only for the intrusion detection on-line [15] but also for medical oriented applications to simulate cancer specific apoptosis [16].

A special feature of the developed approach is that it allows both low-level processing of raw signal [10] and high-level pattern recognition [12, 13]. The Bio nature of the approach together with the speed and accuracy of information processing make immunocomputing uniquely suited for real-life applications. This will probably mean that we will take a further step toward placing more of the intelligent functions on the chip. For this purpose, a hardware emulation of the developing models using digital signal processor (DSP) of the advanced super Harvard architecture (SHARC) has been also proposed.

The obtained results have widely been presented and confirmed by the following ways:

- 16 publications within the project include a book in Springer NY [1] that opens a novel direction of Computer Science and shows a clear way to the world first *immunocomputer* in the nearest years [9, 16];
- EOARD representatives showed interest to this direction during the special meeting on IA within World Congress on Computational Intelligence [2], NASA/DoD Conference on Evolvable Hardware [3], and the 1st Int. Conf. on Artificial Immune Systems sponsored by EOARD [4, 5];

- The feasibility of the developed approach has also been proved through its successful application for such computational intensive problems as learning by at least 40 times faster and recognition by at least 2 times correctly than artificial neural networks and genetic algorithms [12, 13];
- Software emulator of the immunochip and its testing on high dimensional data simulating the task of intrusion detection in a typical US Air Force LAN suggest that the performance of the approach is unachievable for main competitors in the field of Computational Intelligence [14, 15].

Presentation of Project Results

List of published papers

Book:

1. Tarakanov A.O., Skormin V.A., Sokolova S.P. Immunocomputing: Principles and Applications. Springer, New York, 2003.

   Book of the Year Award of the International Institute for Advanced Studies in Systems Research and Cybernetics, Baden-Baden, 2004.

   Review of "Immunocomputing: Principles and Applications by Alexander O. Tarakanov,
   Victor A. Skormin, Svetlana P. Sokolova", Springer-Verlag New York, Inc. 2003
   Source: ACM SIGACT News
   Volume 36, Issue 4 (December 2005)
   Pages: 14-17
   Year of Publication: 2005
   ISSN: 0163-5700
   Author: Wenzhong Zhao, University of New Mexico
   Publisher: ACM Press New York, NY, USA

Papers:

2. Tarakanov A., Skormin V. Pattern recognition by immunocomputing. World Congress on Computational Intelligence, CEC-2002, Vol. 1, pp. 938-943. Honolulu, Hawaii, May 12-17, 2002.
3. Tarakanov A., Dasgupta D. An immunochip architecture and its emulation. NASA/DoD Conference on Evolvable Hardware EH-2002, pp. 261-265. Alexandria, Virginia, July 15-18, 2002.
4. Tarakanov A., Goncharova L., Gupalova T., Kvachev S., Sukhorukov A. Immunocomputing for bioarrays. The 1st Int. Conf. on Artificial Immune Systems ICARIS-2002, pp. 32-40. University of Kent at Canterbury, UK, September 9-11, 2002.
5. Sokolova S., Sokolova L. Immunocomputing for complex interval objects. 1st Int. Conf. on Artificial Immune Systems ICARIS-2002, pp. 222-230.
6. Tarakanov A., Penev G., Madani K. Formal neuro-immune network. Advances in Soft Computing: Neural Networks and Soft Computing. Physica-Verlag, 2002, pp. 644-649.
7. Tarakanov A.O. Spatial formal immune network. Lecture Notes in Computer Science, Vol. 2723, 2003, pp. 248-249.
8. Melnikov Y., Tarakanov A. Immunocomputing model of intrusion detection. Lecture Notes in Computer Science, Vol. 2776, 2003, pp. 453-456.
9. Goncharova L.B., Melnikov Y., Tarakanov A.O. Biomolecular immunocomputing. Lecture Notes in Computer Science, Vol. 2787, 2003, pp. 102-110.
10. Atreas N.D., Karanikas C.G., Tarakanov A.O. Signal processing by an immune type tree transform. Lecture Notes in Computer Science, Vol. 2787, 2003, pp. 111-119.
11. Sokolova L.A. Index design by immunocomputing. Lecture Notes in Computer Science, Vol. 2787, 2003, pp. 120-127.
12. Tarakanov A.O., Tarakanov Y.A. A comparison of immune and neural computing for two real-life tasks of pattern recognition. Lecture Notes in Computer Science, Vol. 3239, 2004, pp. 236-249.
13. Tarakanov A.O., Tarakanov Y.A.: A comparison of immune and genetic algorithms for two real-life tasks of pattern recognition. International Journal of Unconventional Computing, Vol. 1, Issue 4, 2005, pp. 357-374.
14. Tarakanov A.O., Goncharova L.B., Tarakanov O.A.: A cytokine formal immune network. Lecture Notes in Artificial Intelligence, Vol. 3630, 2005, pp. 510-519.
15. Tarakanov A.O., Kvachev S.V., Sukhorukov A.V.: A formal immune network and its implementation for on-line intrusion detection. Lecture Notes in Computer Science, Vol. 3685, 2005, pp. 394-405.
16. Goncharova L.B., Jacques Y., Martin-Vide C., Tarakanov A.O., Timmis J.I.: Biomolecular immune-computer: theoretical basis and experimental simulator. Lecture Notes in Computer Science, Vol. 3627, 2005, pp. 72-85.

List of presentations at conferences and meetings

1) Immunocomputing: Mathematical Basis and Applications. Presentation of A. Tarakanov at the Joint Conference of AFRL, EOARD, SUNY, and SPIIRAS "Novel Information Technologies and Information Assurance", State University of New York (SUNY) at Binghamton, USA, March 4-7, 2002
2) Presentation of paper [2] by A. Tarakanov at the IEEE World Congress on Computational Intelligence (WCCI'02), Honolulu, HI, USA, May 12-17, 2002
3) Immunocomputing: Mathematical Basis and Applications. Presentation of A. Tarakanov at the Dept. of Computational Mathematics and Cybernetics, Moscow State University M.V. Lomonosov, Russia, October 16, 2002 (same as Presentation at SUNY above)
4) Presentation of paper [4] by A. Tarakanov and paper [5] by L. Sokolova at the 1st International Conference on Artificial Immune Systems (ICARIS'02), University of Kent at Canterbury, UK, September 9-11, 2002
5) Presentation of paper [9] by A. Tarakanov and paper [11] by L. Sokolova at the 2nd International Conference on Artificial Immune Systems (ICARIS'03), Napier University, Edinburgh, UK, August 31 – September 4, 2003
6) Presentation of paper [14] by A. Tarakanov at the VIIIth European Conference on Artificial Life (ECAL'05), Canterbury, UK, September 5-9, 2005
7) Presentation of paper [15] by A. Tarakanov at the 3rd International Workshop Mathematical Methods, Models and Architectures for Computer Networks Security (MMM-ACNS'05), St. Petersburg, Russia, September 25-27, 2005.

Information on patents and copyrights

No patents or copyrights were obtained or may be obtained as a result of the project.

Cooperation with Foreign Collaborator

Exchange of scientific material

According to Work Plan, exchange of scientific material included:

- Quarterly reports of project progress:
  - reports Q-1, Q-2, Q-3, Q-5, Q-6, Q-7, Q-9, Q-10, Q-11, Q-13, Q-14, Q-15.
- Annual technical reports of project progress:
  - 1st year report,
  - 2nd year report,
  - 3rd year report.
- Final reports:
  - project activities report,
  - project technical report,
  - project summary for unrestricted distribution.
- Demo versions of the software:
  - demo version 6.1 of the emulator,
  - demo version 6.4 of the emulator,
  - demo version 6.6 of the emulator,
  - demo version 6.7 of the emulator.
- Published papers (e-copies):
  - papers [2], [4], [8], [12], [14], [16].

Signature of protocols

No protocols were signed.

Research carried out jointly

No research was carried out jointly.

Trips to/from foreign collaborators

There were no trips to/from foreign collaborators.

Workshops, topical meetings organized by the project team

There were no workshops or topical meetings organized by the project team.

Joint attendance to international conferences

There were joint attendances to the following international conferences:

- 1st International Conference on Artificial Immune Systems (ICARIS'02), University of Kent at Canterbury, UK, September 9-11, 2002.
- VIIIth European Conference on Artificial Life (ECAL'05), Canterbury, UK, September 5-9, 2005.
- 3rd International Workshop Mathematical Methods, Models and Architectures for Computer Networks Security (MMM-ACNS'05), St. Petersburg, Russia, September 25-27, 2005.

Technology Implementation Plan

How the project results will be implemented in the future work

A hardware implementation of cFIN is proposed based on digital signal processor (DSP) of super Harvard architecture (SHARC).

Perspectives of future developments of the research/technology developed

Two new projects SHARC and BIOCOMP have been proposed.

SHARC project aims to develop a novel approach to intelligent signal processing based on rigorous mathematical models of immunocomputing (IC). A special feature of the approach is that it allows both low-level processing of raw signal and high-level pattern recognition. The Bio nature of the approach together with the speed and accuracy of information processing make IC uniquely suited for real-life applications. This will probably mean that we will take a further step toward placing more of the intelligent functions on the chip. For this purpose, a hardware emulation of the developing models using digital signal processor (DSP) of the advanced super Harvard architecture (SHARC) is also proposed as a proof of principle.

BIOCOMP project proposes to develop a theoretical basis and experimental simulator of the world first Immune-Computer (I-C) as a new kind of biomolecular computer. This I-C will be able to control a fragment of the natural immune system in an autonomous and intelligent manner. Such control has proved unobtainable with other methods.

Potential commercial application of project results

No commercial application of project results is previewed.

Patents and copy rights

No patents or copyrights were obtained or may be obtained as a result of the project.

Project Manager

A.O. Tarakanov

Director of SPIIRAS

R.M. Yusupov

2 February 2006

Final
Project Technical Report
# of ISTC 2200p

Development of Mathematical Models of Immune Networks Intended for Information Security Assurance

(From 1 February 2002 to 31 January 2006 for 48 months)

Alexander Olegovich Tarakanov
(Project Manager)
St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences

February 2006

This work was supported financially by European Office of Aerospace Research and Development (EOARD) and performed under the contract to the International Science and Technology Center (ISTC), Moscow.

Development of Mathematical Models of Immune Networks Intended for Information Security Assurance

(From 1 February 2002 to 31 January 2006 for 48 months)

Alexander Olegovich Tarakanov
(Project Manager)
St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences *

The objective of this project is to develop a novel approach to information assurance (IA) based on a rigorous mathematical notion of formal immune network (FIN).

A special kind of FIN (cFIN) intended for IA has been developed and implemented in so-called immunochip emulator. This software emulator has been tested on data simulating intrusions in a typical computer network (UCI KDD archive). Training time over a training set of about 51000 network connection records is about 60s (AMD 1.5GHz). Fine-tuning of the emulator reduces the number of storing patterns and thus the recognition time per pattern by 60 times at least. The emulator correctly recognizes all intrusions in the training set by 16ms per record.

The comparison with neural computing and genetic algorithms over other real-life tasks of pattern recognition (in ecology and laser physics) also demonstrates that the performance of FIN (training time and accuracy) is unachievable for other approaches of computational intelligence.

A hardware implementation of FIN has been proposed based on digital signal processor of super Harvard architecture (DSP SHARC).

Keywords (about 10 words): Immunocomputing, Information Assurance, Formal Immune Network, Immunochip

*199178, St. Petersburg, 14-line, 39, Russia
Tel: +7 (812) 328 4450, Fax: +7 (812) 328 0685, E-mail: spiiran@iias.spb.su

The work has been performed by the following Institute and Collaborator:

1. Participated Institute:
Saint-Petersburg Institute for Informatics and Automation (SPIIRAS)
14-line, 39, St. Petersburg, 199178, Russia
Tel: +7 (812) 328 4450, Fax: +7 (812) 328 0685, E-mail: spiiran@iias.spb.su

2. Foreign Collaborator:
European Office of Aerospace Research and Development (EOARD)
223-231 Old Marylebone Road, London NW1 5TH United Kingdom
Roberta J. Hauck, USAF EOARD Program Coordinator
Tel: +44 207 514 4354, Fax: +44 207 514 4960, E-mail: roberta.hauck @london.af.mil
Paul Losiewicz, Ph. D., Chief, Information/C41
Tel: +44 207 514 4474, Fax: +44 207 514 4960, E-mail: paul.losiewicz@london.af.mil

List of contents

Introduction

The objective of this project is to develop a novel approach to information assurance (IA) based on a rigorous mathematical notion of formal immune network (FIN).

A special kind of FIN for IA is expected to be developed with the following capabilities:

- FIN as a system of computational intelligence;
- FIN as an alternative to the wide spread artificial neural networks and intelligent agents;
- FIN as a mathematical and algorithmic basis for hardware implementation of IA issues in a special 'immunochip'.

The scope of work and technical approach of the project are as follows:

- Qualitative description of the biological immune networks from the viewpoint of IA.
- Mathematical description of FIN.
- Mathematical description of the special kind of FIN intended for IA.
- Computer implementation of FIN.
- Computer emulator of the immunochip.
- Proposals on hardware implementation of the immunochip.

The 4th year extension of the project has been proposed. It aims to develop a rigorous mathematical model of immune modulation by cytokines (messenger proteins, which play a central role in regulation of immune response) and a novel notion of cytokine FIN (cFIN) intended for IA applications.

An impact of such extension is twofold:

1) The enhanced performance of the software emulator of the immunochip for intrusion detection;

2) The advanced architecture of the immunochip.

Method

Cytokines (messenger proteins) are a group of biologically active mediator molecules that provide the intercellular interactions within the immune system. They are the central regulators of leukocyte growth and differentiation, being produced by a wide variety of cell types, targeting various cell subsets and exhibiting numerous biological activities.

Up to now more than 100 different human cytokines are identified. An increasing volume of experimental data suggests that cytokines play one of the central roles in the immune regulation as well as in the neuro-immune-endocrine modulation.

Recent developments show that cytokines induce apoptosis (programmed cell death) in cancer cells. The induction of apoptosis is associated with a dose-dependent inhibition of cancer cell division, and this activity has been demonstrated for a wide range of cancer types including bladder, breast, leukemia, melanoma, ovarian and prostate.

Apoptosis is a natural mechanism by which cells "commit suicide" when they have outlived their purpose, become defective, or have aged. Apoptosis prevents cells from accumulating and forming tumors. Understanding of the control of apoptosis in normal and malignant cells will help to improve the diagnosis and treatment of malignancies. The goal of many treatments, including chemotherapies is to induce malignant cells to undergo apoptosis. Current data also suggests that a cytokine may function as a dual-acting cytokine in which its normal physiological functions may be related to specific aspects of the immune system and over-expression culminates in cancer-specific apoptosis.

On the other hand, immunological approach looks rather constructive as a basis for a new kind of computing [1]. In such background, this project develops a rigorous mathematical model of immune network with the cytokine controlled apoptosis and immunization. A software implementation of the model has been applied to the task of intrusion detection in a local area network (LAN).

Mathematical model

Cytokine formal immune network

Definition 1. Cell is a pair *V = (c, P)*, where "cytokine" *c* is natural number $c \in N$, whereas $P = (p_1,...,p_q)$ is a point of *q*-dimensional Euclidian space: $P \in R^q$, and P lies within unit cube: $\max\{|p_1|,...,|p_q|\} \le 1$.

Let distance ("affinity") $d_{ij} = d(V_i,V_j)$ between cells V_i and V_j be as follows:

$$d_{ij} = \max\left\{\left|(p_1)_i - (p_1)_j\right|,...,\left|(p_q)_i - (p_q)_j\right|\right\}. \tag{1}$$

Fix some finite non-empty set of cells ("innate immunity") $W_0 = (V_1,...,V_m)$ with non-zero distance between cells: $d_{ij} \ne 0$, $\forall i, j : i \ne j$.

Definition 2. Cytokine formal immune network (cFIN) is a set of cells: $W \subseteq W_0$.

Definition 3. Cell V_i recognizes cell V_k if the following conditions are satisfied: $c_i = c_k$, $d_{ik} < h$, $d_{ik} < d_{ij}$, $\forall V_j \in W$, $j \ne i$, $k \ne j$, where $h \ge 0$ is given "threshold of affinity".

Let us define the behavior ("maturation") of cFIN by the following two rules.

Rule 1 (Apoptosis). If cell $V_i \in W$ recognizes cell $V_k \in W$ then remove V_i from cFIN.

Rule 2 (Auto-Immunization). If cell $V_k \in W$ is nearest to cell $V_i \in W_0 \setminus W$ among all cells of cFIN: $d_{ik} < d_{ij}$, $\forall V_j \in W$, whereas $c_i \ne c_k$, then add V_i to cFIN.

Let W_A be cFIN as a consequent of application of apoptosis to all cells of W_0. Let W_I be cFIN as a consequence of auto-immunization of all cells of W_A by all cells of W_0. Note that the resulting sets W_A and W_I depend on the ordering of cells in W_0. Further it will be assumed that the ordering is given.

Mathematical properties of cFIN

It is obvious that neither the result of apoptosis W_A nor the result of auto-immunization W_I can overcome W_0 for any innate immunity: $W_A \subseteq W_0$, $W_I \subseteq W_0$, $\forall W_0$. Consider more important and less evident properties of cFIN.

Proposition 1. For any innate immunity W_0 there exists threshold of affinity h_0 such that apoptosis does not change W_0 for any *h* less than h_0: $W_A = W_0$, $\forall h < h_0$.

Let h_0 be minimal distance (1) for any pair of cells of cFIN with the same cytokines:

$$h_0 = \min_{i,j}\{d_{ij}\} : \; c_i = c_j, \; i \ne j .$$

Then, according to Definition 3, none of the cells of cFIN can recognize other cells, because $d_{ij} > h_0$ for any pair of cells V_i and V_j. According to Rule 1, none of the cells can be removed from cFIN for any *h* less than h_0, because $d_{ij} > h$, $\forall h < h_0$, $\forall V_i, V_j \in W_0$. Thus, $W_A = W_0$, $\forall h < h_0$.

Proposition 2. For any innate immunity W_0 there exists threshold of affinity h_1 such that consequence of apoptosis and auto-immunization $W_1 = W_I(h_1)$ provides the minimal number of cells $|W_1|$ for given W_0 and any *h*: $|W_1| \le |W_I(h)|$, $\forall h$, $\forall W_I \subseteq W_0$.

Let h_1 be maximal distance (1) for any pair of cells of cFIN with the same cytokines:

$$h_1 = \max_{i,j}\{d_{ij}\} : \; c_i = c_j, \; i \ne j .$$

Then, according to Definition 3, any cell V_i can recognize the nearest cell V_j if the last one has the same cytokine: $c_i = c_j$. Let W_- be the set of all such cells V_i. Then, according to Rule 1, $|W_A(h_1)| = |W_0| - |W_-|$, and such number of cells after apoptosis is minimal among any *h*: $|W_A(h_1)| \le |W_A(h)|$, $\forall h$. Let W_+ be set of cells, which is added to $W_A(h_1)$ as a consequence of auto-immunization: $W_1 = W_A(h_1) \cup W_+$. It is also evident that W_+ is a subset of W_-: $W_+ \subseteq W_-$, and $|W_+|$ represents a number of "mistakes" of apoptosis when cFIN "kills" some cells, which lead to further recognition errors. Such cells are then "restored" by auto-immunization (Rule 2). Let $W_* = W_- \setminus W_+$ be cells which yield apoptosis without further recognition errors. Then $|W_+|=|W_-|-|W_*|$. On the other hand: $|W_1|=|W_A(h_1)|+|W_+|$. Substitutions of $|W_A(h_1)|$ and $|W_+|$ lead to the following result: $|W_1|=|W_0|-|W_*|$. Thus, $|W_1|\le|W_I(h)|$, which proves Proposition 2.

Application of cFIN to pattern recognition

Let "epitope" ("antigenic determinant") be any point $P=(p_1,...,p_q)$ of *q*-dimensional Euclidian space: $P\in R^q$. Note that any cell of cFIN also contains an epitope, according to Definition 1.

Definition 4. Cell V_i recognizes epitope *P* by assigning him class c_i if the distance $d(V_i,P)$ between the cell and the epitope is minimal among all cells of cFIN: $d(V_i,P)=\min\{d(V_j,P)\}, \forall V_j\in W$.

Let pattern be any *n*-dimensional column-vector $Z=[z_1,...,z_n]'$, where $z_1,...,z_n$ are real values and (') is symbol of matrix transposing. Let pattern recognition be mapping of the pattern to an epitope: $Z\to P\in R^q$, and recognition of the epitope by the class of the nearest cell of cFIN. Let $A_1,...,A_m$ be *n*-dimensional training patterns with known classes $c_1,...,c_m$. Let $A=[A_1,...,A_m]'$ be training matrix of dimension $m\times n$. Consider singular value decomposition (SVD: see, e.g., [1]) of this matrix:

$$A=s_1Y_1X_1'+s_2Y_2X_2'+s_3Y_3X_3'+...+s_rY_rX_r',$$

where *r* is the rank of matrix *A*, s_k are singular values and Y_k, X_k are left and right singular vectors with the following properties: $Y_k'Y_k=1$, $X_k'X_k=1$, $Y_k'Y_i=0$, $X_k'X_i=0$, $i\ne k$, $k=1,...,r$, $s_{k-1}\ge s_k$, $k>1$.

Consider the following mapping of any *n*-dimensional pattern Z to epitope P:

$$p_k=\frac{1}{s_k}Z'X_k,\ k=1,...,q,\ q\le r. \tag{2}$$

Note that formulas (2) can be treated as "binding energies" between "formal proteins" Z ("antigens") and X_k ("antibodies"), according to [1]. Note also, that any epitope obtained by application of formulas (2) to any training pattern lies within unit cube (see Definition 1), according to the above properties of singular vectors.

Software implementation

General description (in a pseudocode) of the cFIN algorithm of pattern recognition is as follows:

```
Training
{
    1st stage training // map data to cFIN ("antigen processing")
    {
        Get training patterns;
        Form training matrix;
        Compute SVD of the training matrix; // Singular Value Decomposition
        Store n singular values // "binding energies"
        Store n right singular vectors; // "antibody-probes"
        Store left singular vectors; // cells of cFIN
    }
    2nd stage training // compress data by cFIN's "maturation"
    { // compute consecutively for all cells of cFIN:
        Apoptosis;
        Auto-Immunization;
    }
}
Recognition
{
    Get pattern; // "antigen"
    Map the pattern to cFIN;
    Find nearest cell of cFIN;
    Assign class of the nearest cell to the pattern;
}
```

This algorithm has been implemented in a version of the immunochip emulator (version 6.7) using Visual C++ with build in assembler code (see below: "__asm{…}") of the cytokine affinity function (1) for three-dimensional (3D) Euclidian space ($q=3$) and OpenGL tools for 3D visualization.

```
for (k=1; k<=rows; k++)
{
    if ((k!=i) && (count[k]) && (kill[k] == false))
    {
        xd = (fX[k] - fX[i]);
        yd = (fY[k] - fY[i]);
        zd = (fZ[k] - fZ[i]);
            __asm
                {
                    push    eax
                    push    ebx
                    push    ecx
                    mov     eax,xd
                    mov     ebx,yd
                    mov     ecx,zd
                    shl     eax,1
                    shr     eax,1
                    shl     ebx,1
                    shr     ebx,1
                    shl     ecx,1
                    shr     ecx,1
                    cmp     eax,ebx
                    jg      m1
                    mov     eax,ebx
                m1:
                    cmp     eax,ecx
                    jg      m2
                    mov     eax,ecx
                m2:
                    cmp     eax,dmin
                    ja      m3
                    mov     dmin,eax
                    mov     ebx,k
                    mov     kmin,ebx
                m3:
                    pop     ecx
                    pop     ebx
                    pop     eax
                }
    }
}
```

Fig.1 shows screenshot of the emulator, where right-hand screen shows the initial population of cFIN after SVD, while left-hand screen shows cFIN after apoptosis and auto-immunization.

Results

This cFIN approach has successfully been developed, implemented, and tested as the software emulator of the immunochip.

Two data files from KDD archive (Bay S.D. The UCI KDD Archive [http://kdd.ics.uci.edu]. Irvine, CA: University of California, Dept. of Information and Computer Science, 1999) have been used to test the emulator:

- File 1: kddcup_data_10_percent_gz.htm (7.7 MB);
- File 2: kddcup_newtestdata_10_percent_unlabeled_gz.htm (44 MB).

File 1 is the training data file. It contains 51608 network connection records. Any record (file string) has the following format, where parameters 2, 3, 4, 42 are symbolic, while other 38 parameters are numerical (real values):

```
1) duration, 2) protocol_type, 3) service, 4) flag, 5) src_bytes,
6) dst_bytes, 7) land, 8) wrong_fragment, 9) urgent, 10) hot,
11) num_failed_logins, 12) logged_in, 13) num_compromised,
14) root_shell, 15) su_attempted, 16) num_root, 17) num_file_creations, 18) num_shells,
19) num_access_files, 20) num_outbound_cmds,
21) is_host_login, 22) is_guest_login, 23) count, 24) srv_count,
25) serror_rate, 26) srv_serror_rate, 27) rerror_rate,
28) srv_rerror_rate, 29) same_srv_rate, 30) diff_srv_rate,
31) srv_diff_host_rate, 32) dst_host_count, 33) dst_host_srv_count,
34) dst_host_same_srv_rate, 35) dst_host_diff_srv_rate,
36) dst_host_same_src_port_rate, 37) dst_host_srv_diff_host_rate,
38) dst_host_serror_rate, 39) dst_host_srv_serror_rate,
40) dst_host_rerror_rate, 41) dst_host_srv_rerror_rate, 42) attack_type.
```

For example, two records (# 1 and # 745) of File 1 are as follows:

```
0,tcp,http,SF,181,5450,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,
0.00,0.00,1.00,0.00,0.00,9,9,1.00,0.00,0.11,0.00,0.00,0.00,0.00,0.00, normal.
```

```
184,tcp,telnet,SF,1511,2957,0,0,0,3,0,1,2,1,0,0,1,0,0,0,0,0,1,1,0.00,
0.00,0.00,0.00,1.00,0.00,0.00,1,3,1.00,0.00,1.00,0.67,0.00,0.00,0.00,
0.00, buffer_overflow.
```

File 1.1 has also been prepared with the same 51608 records of the same format just without the last parameter 42) attack_type.

File 2 contains 311079 records of the same format as in File 1.1.

File 1.1 and File 2 are the test data files.

Note that KDD archive does not indicate the correct types of attack for none of the records of File 2. The only available information on possible attacks is gathered in Tab. 1 (column 'Code' is the emulator's code of attack). Nevertheless, File 2 has been used to test whether the emulator is able to detect unknown intrusions, which had not been presented in the training data of File 1.

The results of training the emulator by File 1 are shown in Fig.1, where right-hand screen represents the initial population of cFIN after SVD (Start cells: $|W_0| = 51608$), while left-hand screen shows cFIN after apoptosis and immunization ($h_1 = 0.5$, $|W_1| = 783$). Total training time (for AMD 1.5GHz) is 62 seconds including 8s for the 1st stage (SVD) and 54 s for the 2nd stage (apoptosis and auto-immunization).

During the recognition of the records of File 1.1 and File 2, the emulator writes test results into the output file in the format: Record # - attack_type. For example, four records (## 744-747) with test results for File 1.1 are as follows (see also Tab. 2):

```
744 - normal.
745 - buffer_overflow. !!!
746 - buffer_overflow. !!!
747 - normal.
```

The emulator also shows on-line projection of any pattern to 3D cFIN (see bold skew cross in both screens) and write the recognition result on the bottom panel (see "Class: back !!!").

Test results in Tab. 2 correspond completely to the correct attack types (parameter 42) of File 1.

Another test has been performed over File 2 to check whether the emulator is able to detect unknown intrusions, which had not been presented in the training data of File 1. The intrusion is treated as unknown if the projection of corresponding pattern to cFIN lies outside of the unit cube (according to Definition 1). The emulator has recognized 13 unknown intrusions as the following records ## of File 2:

```
417, 12674, 97891, 139795, 170498, 176201, 177958, 232570, 236975, 296561, 296657, 96796, 297658.
```

According to Tab. 1, any unknown intrusion can correspond to one of the following types of attack that had not been presented in the training data:

```
apache2, guess_passwd, multihop, named, saint, sendmail, snmpgetattack, udpstorm, xlock, xsnoop.
```

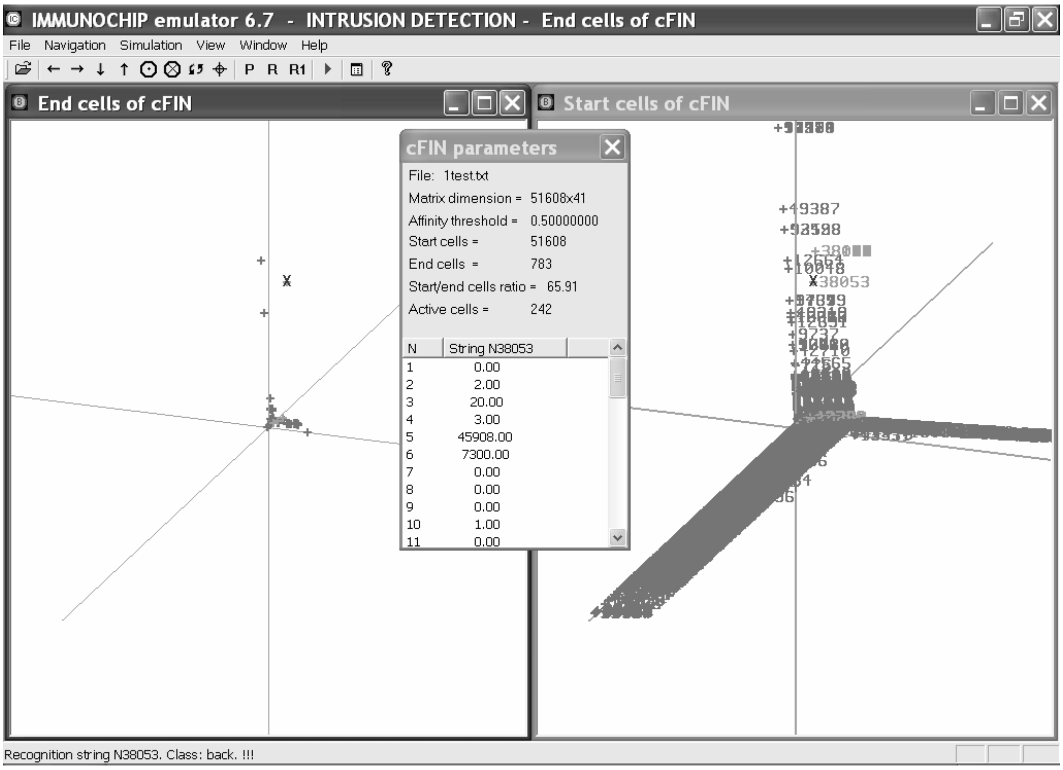
The recognition time per record is 15.7 ms for both tests of File 1.1 and File 2. This time includes not only computations but mainly reading the record from test file, visualization of the recognition result (cFIN's projection of the pattern) in both screens of the emulator and writing the result into output file.

Table 1. Attack types

Code	Attack type	File 1	File 2	Code	Attack type	File 1	File 2
0	normal	+	+				
1	apache2		+	16	pod	+	+
2	back	+		17	portsweep	+	+
3	buffer_overflow	+	+	18	rootkit	+	
4	ftp_write			19	saint		+
5	guess_passwd		+	20	satan	+	
6	imap			21	sendmail		+
7	ipsweep	+	+	22	smurf	+	
8	land	+		23	snmpgetattack		+
9	loadmodule			24	spy		
10	multihop		+	25	teardrop	+	
11	named		+	26	udpstorm		+
12	neptune	+		27	warezclient		
13	nmap			28	warezmaster		
14	perl			29	xlock		+
15	phf	+	+	30	xsnoop		+

Table 2. Test results for File 1.1

Records ##	attack_type	Records ##	attack_type
745-746	Buffer_overflow	38036-38051	ipsweep
3095-7373	Smurf	38052-38151	back
9520-9523	Buffer_overflow	38302-38311	ipsweep
9590-9591	rootkit	42498-42519	ipsweep
9928-10007	neptune	42548-42567	ipsweep
10072	Satan	42593-42594	ipsweep
10320	phf	42706-42708	ipsweep
13340-13519	portsweep	42730-42761	ipsweep
13569	land	42762-42770	buffer_overflow
13845-13864	pod	42771-42772	land
16326-16327	pod	42773-43385	neptune
17446-37902	neptune	44451-44470	neptune
37929-37939	ipsweep	44800-48452	smurf
37959-37963	ipsweep	48453-48552	teadrop
38005-38012	ipsweep	All other	normal


IMMUNOCHIP emulator 6.7 - INTRUSION DETECTION - End cells of cFIN
File Navigation Simulation View Window Help
P R R1
End cells of cFIN
Start cells of cFIN
cFIN parameters
X38053
+19387

File: 1test.txt
Matrix dimension = 51608x41
Affinity threshold = 0.50000000
Start cells = 51608
End cells = 783
Start/end cells ratio = 65.91
Active cells = 242

N	String N38053
1	0.00
2	2.00
3	20.00
4	3.00
5	45908.00
6	7300.00
7	0.00
8	0.00
9	0.00
10	1.00
11	0.00

Recognition string N38053. Class: back. !!!

Fig. 1. Intrusion detection by cFIN: "Antigen" (String 38053 of File 1.1) is mapped to cFIN (skew cross) and recognized by the "cytokine" type of the nearest cell of cFIN (Class: back !!!)

Conclusion

The obtained results suggest that training time and accuracy of the model are beyond the possibilities of artificial neural networks and genetic algorithms [12-15].

According to test results, cFIN reduces the storing patterns by 65.9 times using apoptosis and auto-immunization without any loss of accuracy of recognition. Although this increases the training time (from 8 seconds to 1 minute for AMD 1.5 GHz), nevertheless, more important is the decrease of the recognition time at least by 60 times per pattern by decreasing number of the stored cells of cFIN to be compared with recognizing pattern.

It is worth noting that so good performance of cFIN (error-free recognition with rather low training time) on the data of real-life dimension looks unobtainable for main competitors in the field of computational intelligence like artificial neural networks (ANN) and genetic algorithms (GA). According to the comparison in [12] and [13], cFIN trains by at least 40 times faster and recognizes by at least 2 times correctly than ANN and GA on the tasks of environmental monitoring and laser physics. These tasks have rather low dimension: 17×23×6 for ecological atlas and 19×5 for laser diode. Such drawbacks of ANN and GA become especially inadmissible for the task of intrusion detection with rather high dimension 51608×41 and more.

It is also worth noting that cFIN differs essentially from the negative selection algorithm (NSA). Actually, NSA aims to provide a set of detectors for self-nonself discrimination, whereas cFIN guarantees a minimal set of "cells" for the correct recognition of any number of classes based on "cytokines". Apparently, this makes cFIN advantageous not only for the intrusion detection on-line [15] but also for medical oriented applications to simulate cancer specific apoptosis [16].

A special feature of the developed approach is that it allows both low-level processing of raw signal [10] and high-level pattern recognition [12, 13]. The Bio nature of the approach together with the speed and accuracy of information processing make immunocomputing uniquely suited for real-life applications. This will probably mean that we will take a further step toward placing more of the intelligent functions on the chip. For this purpose, a hardware emulation of the developing models using digital signal processor (DSP) of the advanced super Harvard architecture (SHARC) has been also proposed.

The obtained results have widely been presented and confirmed by the following ways:

- 16 publications within the project include a book in Springer NY [1] that opens a novel direction of Computer Science and shows a clear way to the world first *immunocomputer* in the nearest years [9, 16];
- EOARD representatives showed interest to this direction during the special meeting on IA within World Congress on Computational Intelligence [2], NASA/DoD Conference on Evolvable Hardware [3], and the 1st Int. Conf. on Artificial Immune Systems sponsored by EOARD [4, 5];
- The feasibility of the developed approach has also been proved through its successful application for such computational intensive problems as learning by at least 40 times faster and recognition by at least 2 times correctly than artificial neural networks and genetic algorithms [12, 13];
- Software emulator of the immunochip and its testing on high dimensional data simulating the task of intrusion detection in a typical US Air Force LAN suggest that the performance of the approach is unachievable for main competitors in the field of Computational Intelligence [14, 15].

List of published papers with abstracts

Book:

1. Tarakanov A.O., Skormin V.A., Sokolova S.P. Immunocomputing: Principles and Applications. Springer, New York, 2003.

This book introduces *immunocomputing* (IC) as a new computing approach that replicates the principles of information processing by proteins and immune networks. It establishes a rigorous mathematical basis for IC, consistent with recent findings in immunology, and it presents various applications of IC to specific computationally intensive real-life problems. The hardware implementation aspects of the IC concept in an *immunocomputer* as a new kind of computing medium and its potential connections with modern biological microchips (biochips) and future biomolecular computers (biocomputers) are also discussed.

Topics and features:

- Establishes a strong mathematical basis for IC, consistent with recent findings in immunology and biochip development;
- Provides a rigorous approach to a hardware implementation of artificial immune systems in 'immunochips';
- Integrates key aspects of pattem recognition, language representation, and knowledge-based reasoning;
- Examines key applications – protein modeling, space navigation, information security protection, infection control, and ecology;
- Outlines IC's potential for creating biological microchips and biomolecular computers.

This thorough introduction to immunocomputing is a valuable resource for experts in computer science, artificial intelligence, and biomolecular computing who would like to explore the principles of IC, as well as for immunologists seeking to further quantify their research. It will also assist multidisciplinary researchers in mutually enhancing computer science and immunology methods.

Book of the Year Award of the International Institute for Advanced Studies in Systems Research and Cybernetics, Baden-Baden, 2004.

Review of "Immunocomputing: Principles and Applications by Alexander O. Tarakanov,
Victor A. Skormin, Svetlana P. Sokolova", Springer-Verlag New York, Inc. 2003
Source: ACM SIGACT News
Volume 36, Issue 4 (December 2005)

Pages: 14-17
Year of Publication: 2005
ISSN: 0163-5700
Author: Wenzhong Zhao, University of New Mexico
Publisher: ACM Press New York, NY, USA

Papers:

2. Tarakanov A., Skormin V. Pattern recognition by immunocomputing. World Congress on Computational Intelligence, CEC-2002, Vol. 1, pp. 938-943. Honolulu, Hawaii, May 12-17, 2002.

The authors had developed a rigorous mathematical approach, describing operation of the immune system based on the models of proteins and immune networks. This approach, Immunocomputing, has been proposed as a computational basis for Artificial Immune Systems. A further development of *Immunocomputing* and its application to pattern recognition is considered herein. It is shown that intrusion detection in computer networks presents a possible implementation of Immunocomputing.

3. Tarakanov A., Dasgupta D. An immunochip architecture and its emulation. NASA/DoD Conference on Evolvable Hardware EH-2002, pp. 261-265. Alexandria, Virginia, July 15-18, 2002.

The paper proposes an architecture for building immunochips and provides a mathematical framework in describing some of its operations using the concepts of proteins and immune networks. This approach is considered as the computational basis of an “immunochip”, and this paper describes its implementation procedure. The proposed immunochip is emulated in software and evaluated with the problem of detecting of dangerous ballistic situations in near-Earth space.

4. Tarakanov A., Goncharova L., Gupalova T., Kvachev S., Sukhorukov A. Immunocomputing for bioarrays. The 1st Int. Conf. on Artificial Immune Systems ICARIS-2002, pp. 32-40. University of Kent at Canterbury, UK, September 9-11, 2002.

This paper presents results of application of our immunocomputing method to immune diagnostic arrays. The method detects bound complexes of immunoglobulin G (IgG) with protein G (pG), and recognizes the concentration of IgG as the result of IgG-pG interactions at each location of a bioarray. This model system has been developed as a prototype of a protein biochip for immunoassay-based diagnostics, where bioarray is a macro-variant of the biochip microarray, while the software is a core of the biochip reader and controller.

5. Sokolova S., Sokolova L. Immunocomputing for complex interval objects. 1st Int. Conf. on Artificial Immune Systems ICARIS-2002, pp. 222-230.

This paper provides a further development of the Immunocomputing (IC) approach to the class of complex objects with parameter uncertainty of the interval type. By using the rules and nomenclature of interval mathematics the singular value decomposition (SVD) of interval matrices, procedures for supervised learning, unsupervised learning, classification and presentation of the results of research in IC shape space have been further developed. This paper includes examples of Specific Interval Artificial Immune Systems for Surveillance of the Plague and Security Systems.

6. Tarakanov A., Penev G., Madani K. Formal neuro-immune network. Advances in Soft Computing: Neural Networks and Soft Computing. Physica-Verlag, 2002, pp. 644-649.

The paper presents an attempt to introduce a new formal notion of Neuro-Immune Network (NIN) based on a rigorous mathematical basis. This notion is inspired by a biological phenomenon of reciprocal impact between the neural and immune systems. The paper considers examples of NIN including a possible application to intrusion detection in computer networks.

7. Tarakanov A.O. Spatial formal immune network. Lecture Notes in Computer Science, Vol. 2723, 2003, pp. 248-249.

A notion of Spatial Formal Immune Network is proposed for pattern recognition applications.

8. Melnikov Y., Tarakanov A. Immunocomputing model of intrusion detection. Lecture Notes in Computer Science, Vol. 2776, 2003, pp. 453-456.

The paper proposes an immunocomputing model of intrusion detection based on a mathematical notion of formal immune network. An application example is provided using software emulator of an immunochip.

9. Goncharova L.B., Melnikov Y., Tarakanov A.O. Biomolecular immunocomputing. Lecture Notes in Computer Science, Vol. 2787, 2003, pp. 102-110.

The paper proposes a new application of biomolecular computing to processing of ex vivo fragment of computer controlled immune system. Our approach involves two basic components: the immunocomputing computational paradigm and a protein biochip to provide a direct interface between the immune system and the computer hardware.

10. Atreas N.D., Karanikas C.G., Tarakanov A.O. Signal processing by an immune type tree transform. Lecture Notes in Computer Science, Vol. 2787, 2003, pp. 111-119.

The paper makes an attempt to introduce a new approach for detection of local singularities in signals, including one-dimensional time series and two-dimensional images. Inspired by a mode of antigen processing in the immune system, our approach is based on the rigorous mathematical methods of Discrete Tree Transform (DTT) and Singular Value Decomposition (SVD). The approach has successfully been applied to detect local singularities in human electrocardiogram (ECG), as well as to enhance the detection of bound complexes of human immunoglobulin in biochip-like bio-membranes.

11. Sokolova L.A. Index design by immunocomputing. Lecture Notes in Computer Science, Vol. 2787, 2003, pp. 120-127.

This paper presents the concept of applying indices, data fusion and mathematical models using immunocomputing approach. The application of indices by immunocomputing can reduce large quantities of variable data relating to a complex interacting dynamic system, into a single general value or index that represents all of those factors (data fusion) to achieve a solution to a practical problem. To illustrate the concept, this article provides examples of mathematical models showing the identification of intrusions into computer networks and the occurrence of plague in Kazakhstan.

12. Tarakanov A.O., Tarakanov Y.A. A comparison of immune and neural computing for two real-life tasks of pattern recognition. Lecture Notes in Computer Science, Vol. 3239, 2004, pp. 236-249.

This paper compares a new Immunocomputing (IC) approach with Artificial Neural Networks (ANN). We compare an IC algorithm of pattern recognition with Error Back Propagation (EBP) network. The comparison includes two real-life tasks of environmental monitoring and laser physics.

13. Tarakanov A.O., Tarakanov Y.A.: A comparison of immune and genetic algorithms for two real-life tasks of pattern recognition. International Journal of Unconventional Computing, Vol. 1, Issue 4, 2005, pp. 357-374.

This paper compares a new Immunocomputing (IC) approach with Genetic Algorithms (GAs). We compare an IC algorithm of pattern recognition with a basic GA. The comparison includes supervised learning over two real-life tasks of environmental monitoring and laser physics.

14. Tarakanov A.O., Goncharova L.B., Tarakanov O.A.: A cytokine formal immune network. Lecture Notes in Artificial Intelligence, Vol. 3630, 2005, pp. 510-519.

This paper develops a mathematical model of immune network controlled by cytokines. A software implementation of the model has been applied to intrusion detection in computer network. The obtained results suggest that the performance of the model is unachievable for another approaches of computational intelligence.

15. Tarakanov A.O., Kvachev S.V., Sukhorukov A.V.: A formal immune network and its implementation for on-line intrusion detection. Lecture Notes in Computer Science, Vol. 3685, 2005, pp. 394-405.

This paper presents a mathematical model of immune network specified for real-time intrusion detection. A software implementation of the model has been tested on data simulating a typical US Air Force local area network (LAN). The obtained results suggest that the performance of the model is unachievable for other approaches of computational intelligence. A hardware implementation of the model is proposed based on digital signal processor (DSP) of super Harvard architecture (SHARC).

16. Goncharova L.B., Jacques Y., Martin-Vide C., Tarakanov A.O., Timmis J.I.: Biomolecular immune-computer: theoretical basis and experimental simulator. Lecture Notes in Computer Science, Vol. 3627, 2005, pp. 72-85.

We propose to develop a theoretical basis and experimental simulator of the first Immune-Computer (IC) as a new kind of biomolecular computer. This IC will be able to control a fragment of the natural immune system in an autonomous and intelligent manner. Such control has proved unobtainable with other methods.

List of presentations at conferences and meetings with abstracts

1) Immunocomputing: Mathematical Basis and Applications. Presentation of A. Tarakanov at the Joint Conference of AFRL, EOARD, SUNY, and SPIIRAS "Novel Information Technologies and Information Assurance", State University of New York (SUNY) at Binghamton, USA, March 4-7, 2002

This talk presented our results in developing a rigorous mathematical basis of a novel approach to computing, immunocomputing, and its applications to solve specific real world problems. Immunocomputing was inspired by the biological principles of information processing by proteins and immune networks. We introduced new mathematical abstractions of Formal Protein and Formal Immune Network (FIN). We provided a rigorous proof that a FIN was able to learn, to recognize, to solve problems and to represent languages based on the theory of linguistic valence. We presented some applied results, such as computing of ecological atlases, monitoring of most dangerous infections, detecting critical situations in near Earth space, information security, etc. These results allowed us to speak about the hardware implementation of the immunocomputing in so-called immunochips. Our project aimed at the development of a biochip (biological microchip) has been also discussed. The presentation was illustrated by several versions of our software emulator of the immunochip.

2) Presentation of paper [2] by A. Tarakanov at the IEEE World Congress on Computational Intelligence (WCCI'02), Honolulu, HI, USA, May 12-17, 2002

3) Immunocomputing: Mathematical Basis and Applications. Presentation of A. Tarakanov at the Dept. of Computational Mathematics and Cybernetics, Moscow State University M.V. Lomonosov, Russia, October 16, 2002 (same as Presentation at SUNY above)

4) Presentation of paper [4] by A. Tarakanov and paper [5] by L. Sokolova at the 1st International Conference on Artificial Immune Systems (ICARIS'02), University of Kent at Canterbury, UK, September 9-11, 2002

5) Presentation of paper [9] by A. Tarakanov and paper [11] by L. Sokolova at the 2nd International Conference on Artificial Immune Systems (ICARIS'03), Napier University, Edinburgh, UK, August 31 – September 4, 2003

6) Presentation of paper [14] by A. Tarakanov at the VIIIth European Conference on Artificial Life (ECAL'05), Canterbury, UK, September 5-9, 2005

7) Presentation of paper [15] by A. Tarakanov at the 3rd International Workshop Mathematical Methods, Models and Architectures for Computer Networks Security (MMM-ACNS'05), St. Petersburg, Russia, September 25-27, 2005.

Information on patents and copyrights

No patents or copyrights were obtained or may be obtained as a result of the project.

Project Manager

A.O. Tarakanov

Director of SPIIRAS

R.M. Yusupov

2 February 2006

Final
Project Summary Report
# of ISTC 2200p

Development of Mathematical Models of Immune Networks Intended for Information Security Assurance

(From 1 February 2002 to 31 January 2006 for 48 months)

Alexander Olegovich Tarakanov
(Project Manager)
St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences

February 2006

This work was supported financially by European Office of Aerospace Research and Development (EOARD) and performed under the contract to the International Science and Technology Center (ISTC), Moscow.

Development of Mathematical Models of Immune Networks Intended for Information Security Assurance

(From 1 February 2002 to 31 January 2006 for 48 months)

Alexander Olegovich Tarakanov
(Project Manager)
St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences *

The objective of this project is to develop a novel approach to information assurance (IA) based on a rigorous mathematical notion of formal immune network (FIN).

A special kind of FIN (cFIN) intended for IA has been developed and implemented in so-called immunochip emulator. This software emulator has been tested on data simulating intrusions in a typical computer network (UCI KDD archive). Training time over a training set of about 51000 network connection records is about 60s (AMD 1.5GHz). Fine-tuning of the emulator reduces the number of storing patterns and thus the recognition time per pattern by 60 times at least. The emulator correctly recognizes all intrusions in the training set by 16ms per record.

The comparison with neural computing and genetic algorithms over other real-life tasks of pattern recognition (in ecology and laser physics) also demonstrates that the performance of FIN (training time and accuracy) is unachievable for other approaches of computational intelligence.

A hardware implementation of FIN has been proposed based on digital signal processor of super Harvard architecture (DSP SHARC).

Keywords (about 10 words): Immunocomputing, Information Assurance, Formal Immune Network, Immunochip

*199178, St. Petersburg, 14-line, 39, Russia
Tel: +7 (812) 328 4450, Fax: +7 (812) 328 0685, E-mail: spiiran@iias.spb.su

**The work has been performed by the following
Institute and Collaborator:**

1. Participated Institute:
Saint-Petersburg Institute for Informatics and Automation (SPIIRAS)
14-line, 39, St. Petersburg, 199178, Russia
Tel: +7 (812) 328 4450, Fax: +7 (812) 328 0685, E-mail: spiiran@iias.spb.su

2. Foreign Collaborator:
European Office of Aerospace Research and Development (EOARD)
223-231 Old Marylebone Road, London NW1 5TH United Kingdom
Roberta J. Hauck, USAF EOARD Program Coordinator
Tel: +44 207 514 4354, Fax: +44 207 514 4960, E-mail: roberta.hauck @london.af.mil
Paul Losiewicz, Ph. D., Chief, Information/C41
Tel: +44 207 514 4474, Fax: +44 207 514 4960, E-mail: paul.losiewicz@london.af.mil

Summary of the project

The objective of this project is to develop a novel approach to information assurance (IA) based on a rigorous mathematical notion of formal immune network (FIN).

With its highly parallel, adaptive processes and capacity for efficiently recognizing and classifying tasks, the immune system provides an excellent information-processing model for designing a powerful computing device. Harnessing these natural-computing methods is instrumental in solving computationally intensive, complex problems, including IA.

The mathematical formalization of these capabilities forms the basis of the new computational approach *immunocomputing* (IC) that includes the notion of FIN.

The scope of work and technical approach of the project are as follows:

- Qualitative description of the biological immune networks from the viewpoint of IA.
- Mathematical description of FIN.
- Mathematical description of the special kind of FIN intended for IA.
- Computer implementation of FIN.
- Computer emulator of the immunochip.
- Proposals on hardware implementation of the immunochip.

The 4th year extension of the project has been proposed. It aims to develop a rigorous mathematical model of immune modulation by cytokines (messenger proteins, which play a central role in regulation of immune response) and a novel notion of cytokine FIN (cFIN) intended for IA applications.

An impact of such extension is twofold:

1) The enhanced performance of the software emulator of the immunochip for intrusion detection;

2) The advanced architecture of the immunochip.

EUROGRAM #05-01 (Jan-Feb 05)
European Office of Aerospace Research and Development
(page 2):

"Grant Awarded: Dr. Alexander Tarakanov was awarded a cost extension to “Development of Mathematical Models of Immune Networks Intended for Information Security Assurance” to expand his immuno-computing Formal Immune Network (FIN) model to include a cytokine-modeled FIN (cFIN). This is the last software emulation task required prior to actual implementation of the emulator in hardware. Dr. Tarakanov co-authored "Immunocomputing: Principles and Applications", winner of book-of-the-year recognition by the International Institute for Advanced Studies in Systems Research and Cybernetics. AFRL lab evaluator is Mr. Joseph Giordano AFRL/IFGB."

General description (in a pseudocode) of the cFIN approach to pattern recognition is as follows:

```
Training
{
    1st stage training // map data to cFIN ("antigen processing")
    {
        Get training patterns;
        Form training matrix;
        Compute SVD of the training matrix; // Singular Value Decomposition
        Store n singular values // "binding energies"
        Store n right singular vectors; // "antibody-probes"
        Store left singular vectors; // cells of cFIN
    }
    2nd stage training // compress data by cFIN's "maturation"
    { // compute consecutively for all cells of cFIN:
        Apoptosis;
        Auto-Immunization;
    }
}
Recognition
{
    Get pattern; // "antigen"
    Map the pattern to cFIN;
    Find nearest cell of cFIN;
    Assign class of the nearest cell to the pattern;
}
```

This cFIN approach has successfully been developed, implemented, and tested as the software emulator of the immunochip.

Two data files from KDD archive (Bay S.D. The UCI KDD Archive [http://kdd.ics.uci.edu]. Irvine, CA: University of California, Dept. of Information and Computer Science, 1999) have been used to test the emulator:

- File 1: kddcup_data_10_percent_gz.htm (7.7 MB);
- File 2: kddcup_newtestdata_10_percent_unlabeled_gz.htm (44 MB).

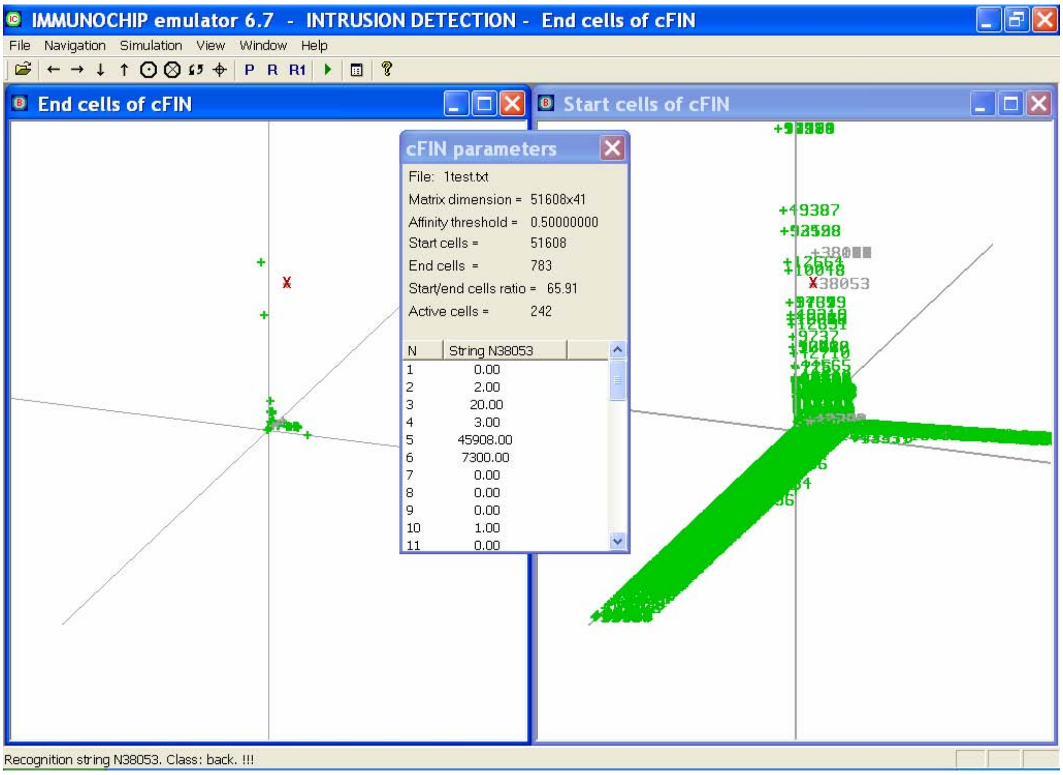
File 1 is the training data file. It contains 51608 network connection records.

File 1.1 has also been prepared with the same 51608 records of the same format just without the last parameter ('attack_type'). File 2 contains 311079 records of the same format as in File 1.1. File 1.1 and File 2 are the test data files.

The results of training the emulator by File 1 are shown in Fig.1, where right-hand screen represents the initial population of cFIN after SVD, while left-hand screen shows cFIN after apoptosis and immunization. Total training time (for AMD Athlon 1.5GHz) is 62 seconds including 8s for the 1st stage (SVD) and 54s for the 2nd stage (apoptosis and auto-immunization).

Test results for File 1.1 correspond completely to the correct attack types of File 1.

The emulator has recognized 13 unknown intrusions in File 2.



File: 1test.txt
Matrix dimension = 51608x41
Affinity threshold = 0.50000000
Start cells = 51608
End cells = 783
Start/end cells ratio = 65.91
Active cells = 242

N	String N38053
1	0.00
2	2.00
3	20.00
4	3.00
5	45908.00
6	7300.00
7	0.00
8	0.00
9	0.00
10	1.00
11	0.00

Recognition string N38053. Class: back. !!!

Fig. 1. Intrusion detection by cFIN: "Antigen" (String 38053 of File 1.1) is mapped to cFIN (skew cross) and recognized by the "cytokine" type of the nearest cell of cFIN (Class: back !!!)

The obtained results have widely been presented and confirmed by the following ways:

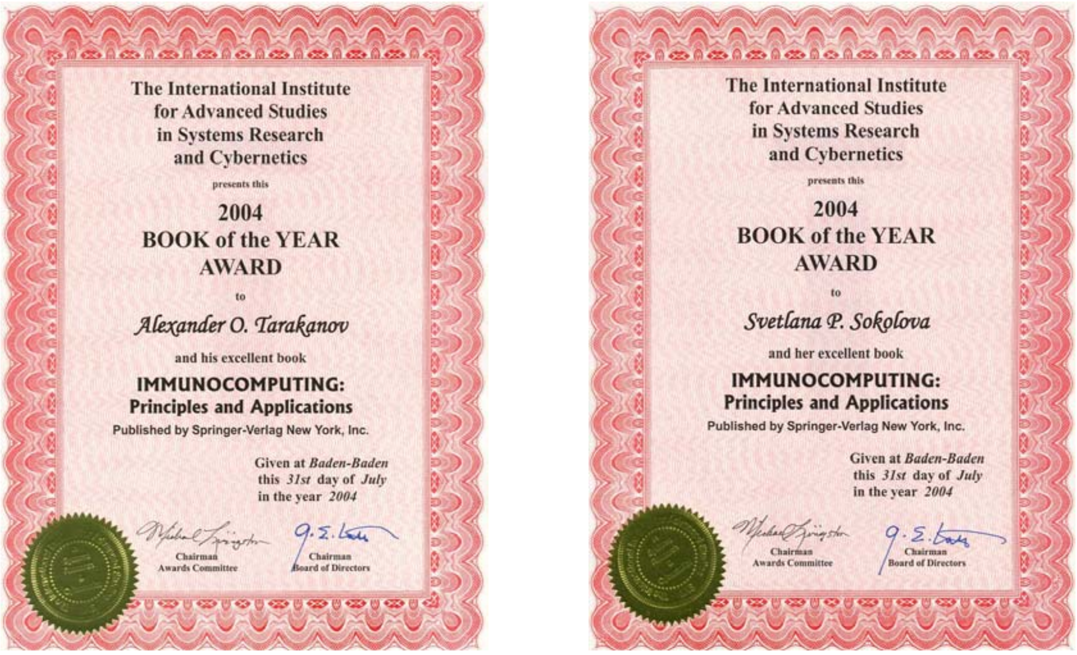
- 16 publications within the project include a book in Springer NY [1] that opens a novel direction of Computer Science and shows a clear way to the world first *immunocomputer* in the nearest years [9, 16];
- EOARD representatives showed interest to this direction during the special meeting on IA within World Congress on Computational Intelligence [2], NASA/DoD Conference on Evolvable Hardware [3], and the 1st Int. Conf. on Artificial Immune Systems sponsored by EOARD [4, 5];
- The feasibility of the developed approach has also been proved through its successful application for such computational intensive problems as learning by at least 40 times faster and recognition by at least 2 times correctly than artificial neural networks and genetic algorithms [12, 13];
- Software emulator of the immunochip and its testing on high dimensional data simulating the task of intrusion detection in a typical US Air Force LAN suggest that the performance of the approach is unachievable for main competitors in the field of Computational Intelligence [14, 15].

List of publications

Book:

1. Tarakanov A.O., Skormin V.A., Sokolova S.P. Immunocomputing: Principles and Applications. Springer, New York, 2003.

   Book of the Year Award of the International Institute for Advanced Studies in Systems Research and Cybernetics, Baden-Baden, 2004 (Fig. 2).



The International Institute
for Advanced Studies
in Systems Research
and Cybernetics

presents this

2004
BOOK of the YEAR
AWARD

to

Alexander O. Tarakanov

and his excellent book

IMMUNOCOMPUTING:
Principles and Applications
Published by Springer-Verlag New York, Inc.

Given at *Baden-Baden*
this *31st* day of *July*
in the year *2004*

Chairman
Awards Committee

Chairman
Board of Directors

The International Institute
for Advanced Studies
in Systems Research
and Cybernetics

presents this

2004
BOOK of the YEAR
AWARD

to

Svetlana P. Sokolova

and her excellent book

IMMUNOCOMPUTING:
Principles and Applications
Published by Springer-Verlag New York, Inc.

Given at *Baden-Baden*
this *31st* day of *July*
in the year *2004*

Chairman
Awards Committee

Chairman
Board of Directors

Fig. 2. Book of the Year Award

Review of "Immunocomputing: Principles and Applications by Alexander O. Tarakanov,
Victor A. Skormin, Svetlana P. Sokolova", Springer-Verlag New York, Inc. 2003
Source: ACM SIGACT News
Volume 36, Issue 4 (December 2005)
Pages: 14-17
Year of Publication: 2005
ISSN: 0163-5700
Author: Wenzhong Zhao, University of New Mexico
Publisher: ACM Press New York, NY, USA

Papers:

2. Tarakanov A., Skormin V. Pattern recognition by immunocomputing. World Congress on Computational Intelligence, CEC-2002, Vol. 1, pp. 938-943. Honolulu, Hawaii, May 12-17, 2002.
3. Tarakanov A., Dasgupta D. An immunochip architecture and its emulation. NASA/DoD Conference on Evolvable Hardware EH-2002, pp. 261-265. Alexandria, Virginia, July 15-18, 2002.
4. Tarakanov A., Goncharova L., Gupalova T., Kvachev S., Sukhorukov A. Immunocomputing for bioarrays. The 1st Int. Conf. on Artificial Immune Systems ICARIS-2002, pp. 32-40. University of Kent at Canterbury, UK, September 9-11, 2002.
5. Sokolova S., Sokolova L. Immunocomputing for complex interval objects. 1st Int. Conf. on Artificial Immune Systems ICARIS-2002, pp. 222-230.
6. Tarakanov A., Penev G., Madani K. Formal neuro-immune network. Advances in Soft Computing: Neural Networks and Soft Computing. Physica-Verlag, 2002, pp. 644-649.
7. Tarakanov A.O. Spatial formal immune network. Lecture Notes in Computer Science, Vol. 2723, 2003, pp. 248-249.
8. Melnikov Y., Tarakanov A. Immunocomputing model of intrusion detection. Lecture Notes in Computer Science, Vol. 2776, 2003, pp. 453-456.
9. Goncharova L.B., Melnikov Y., Tarakanov A.O. Biomolecular immunocomputing. Lecture Notes in Computer Science, Vol. 2787, 2003, pp. 102-110.

10. Atreas N.D., Karanikas C.G., Tarakanov A.O. Signal processing by an immune type tree transform. Lecture Notes in Computer Science, Vol. 2787, 2003, pp. 111-119.
11. Sokolova L.A. Index design by immunocomputing. Lecture Notes in Computer Science, Vol. 2787, 2003, pp. 120-127.
12. Tarakanov A.O., Tarakanov Y.A. A comparison of immune and neural computing for two real-life tasks of pattern recognition. Lecture Notes in Computer Science, Vol. 3239, 2004, pp. 236-249.
13. Tarakanov A.O., Tarakanov Y.A.: A comparison of immune and genetic algorithms for two real-life tasks of pattern recognition. International Journal of Unconventional Computing, Vol. 1, Issue 4, 2005, pp. 357-374.
14. Tarakanov A.O., Goncharova L.B., Tarakanov O.A.: A cytokine formal immune network. Lecture Notes in Artificial Intelligence, Vol. 3630, 2005, pp. 510-519.
15. Tarakanov A.O., Kvachev S.V., Sukhorukov A.V.: A formal immune network and its implementation for on-line intrusion detection. Lecture Notes in Computer Science, Vol. 3685, 2005, pp. 394-405.
16. Goncharova L.B., Jacques Y., Martin-Vide C., Tarakanov A.O., Timmis J.I.: Biomolecular immune-computer: theoretical basis and experimental simulator. Lecture Notes in Computer Science, Vol. 3627, 2005, pp. 72-85.

Project Manager

A.O. Tarakanov

Director of SPIIRAS

R.M. Yusupov

2 February 2006